

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE INFORMÁTICA
CURSO DE CIÊNCIA DA COMPUTAÇÃO

GUILHERME AUGUSTO DIAS

**Análise de Sentimento em Artefatos de
Software**

Monografia apresentada como requisito parcial
para a obtenção do grau de Bacharel em Ciência
da Computação

Orientador: Prof^a. Dr^a. Karin Becker

Porto Alegre
2018

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

Reitor: Prof. Rui Vicente Oppermann

Vice-Reitora: Prof^a. Jane Fraga Tutikian

Pró-Reitor de Graduação: Prof. Wladimir Pinheiro do Nascimento

Diretora do Instituto de Informática: Prof^a. Carla Maria Dal Sasso Freitas

Coordenador do Curso de Ciência de Computação: Prof. Sérgio Luis Cechin

Bibliotecária-chefe do Instituto de Informática: Beatriz Regina Bastos Haro

AGRADECIMENTOS

Gostaria de agradecer a minha mãe, a meu pai e a minha irmã que prestaram todo tipo de apoio imaginável ao longo minha jornada acadêmica. A Caroline Moers pela parceria, compreensão e motivação que me trouxe ao longo dos meses de desenvolvimento deste trabalho. À professora Karin Becker, cuja paciência e dedicação na orientação viabilizaram este trabalho. Aos meus colegas de curso por compartilhar os desafios e enriquecer minhas experiências e aprendizados. Aos meus colegas de time pela compreensão e apoio ao longo das semanas finais do desenvolvimento deste trabalho.

RESUMO

Sentimento é um componente que permeia o cotidiano de organizações, tendo impacto em diversos aspectos que vão da tomada de decisão à performance de funcionários. Neste trabalho verifica-se a existência de relações entre a assertividade na descrição de cartões Jira, medida na forma de emoções expressas nos mesmos, e métricas de produtividade, qualidade e sentimento de pessoas envolvidas na resolução do cartão. Além disso verifica-se também a relação do sentimento expresso e tipo de trabalho representado pelo cartão.

Palavras-chave: Sentimento. Jira.

ABSTRACT

Sentiment is a pervasive component in the daily operations of organizations, having impact in several aspects, ranging from decision making process to employees performance. In this work we verify the existence of relationships between assertiveness in the description of Jira cards, evaluated as the emotions expressed in the text, and productivity metrics, quality metrics and sentiment of people involved in the card resolution. Furthermore, we verify the relation between the sentiment and type of work represented by the Jira card.

Keywords: sentiment, jira.

LISTA DE FIGURAS

Figura 2.1 Exemplo de cartão Jira.....	18
Figura 2.2 Exemplo de quadro Jira	19
Figura 3.1 Fluxo de trabalho da Bay Bridge	23
Figura 4.1 Entrada no histórico de alterações	30
Figura 5.1 Dominância do título vs. tempo de resolução.....	31
Figura 5.2 Dominância da descrição vs. tempo de resolução	32
Figura 5.3 Dominância do título vs. rejeições	33
Figura 5.4 Dominância da descrição vs. rejeições	33
Figura 5.5 Dominância da descrição vs. valência dos comentários	34
Figura 5.6 Dominância da descrição vs. excitação dos comentários	35
Figura 5.7 Dominância da descrição vs. dominância dos comentários	35
Figura 5.8 Valência da descrição vs. tipo de trabalho	36
Figura 5.9 Valência da descrição vs. tipo de trabalho	37
Figura 5.10 Valência da descrição vs. tipo de trabalho	37

LISTA DE TABELAS

Tabela 2.1 Exemplos de avaliação VAD	15
--	----

LISTA DE ABREVIATURAS E SIGLAS

PLN Processamento de Linguagem Natural

VAD Valence-Arousal-Dominance

UFRGS Universidade Federal do Rio Grande do Sul

SUMÁRIO

1 INTRODUÇÃO	10
2 FUNDAMENTAÇÃO TEÓRICA	12
2.1 Análise de Sentimento	12
2.2 Frameworks para Teorias de Emoção	12
2.2.1 Framework Discreto.....	13
2.2.2 Framework Dimensional.....	13
2.3 Técnicas de Análise de Sentimento	13
2.4 Dicionários de Emoção	15
2.5 Processamento de Linguagem Natural	16
2.6 Scrum	16
2.6.1 Papéis	16
2.6.2 Artefatos.....	17
2.6.3 Ciclo de Desenvolvimento Scrum.....	17
2.7 Jira	18
2.8 Trabalhos Relacionados	19
3 CONTEXTO DA ORGANIZAÇÃO BAY BRIDGE	21
3.1 Organização	21
3.2 Processo de Desenvolvimento	22
3.2.1 Backlog do produto.....	22
3.2.2 Ciclos de desenvolvimento	23
3.3 Motivação	24
4 MÉTODO	25
4.1 Análises propostas	25
4.2 Coleta de Dados	26
4.3 Cálculo de Sentimento	27
4.3.1 Amplitude máxima.....	28
4.3.2 Maior distância da média	29
4.4 Tratamento aos Dados	29
4.4.1 Tempo de resolução	30
4.4.2 Número de rejeições	30
5 RESULTADOS	31
5.1 Assertividade vs. Métricas de Produtividade e Qualidade	31
5.1.1 Produtividade	31
5.1.2 Qualidade	32
5.2 Assertividade vs. Sentimento nos Comentários	34
5.3 Tipo de trabalho vs. Sentimento no Cartão	36
6 CONCLUSÃO	39
REFERÊNCIAS	40

1 INTRODUÇÃO

Ao longo dos últimos anos, observamos a tecnologia da informação ser incorporada à base da operação de organizações. Em muitos casos, o desenvolvimento de software tornou-se componente central e decisivo para o sucesso de empresas. Em resposta a necessidades de um mercado cada vez mais competitivo e em constante mudança, diversas metodologias de desenvolvimento de software foram criadas, como XP e Scrum. Tais metodologias buscam prover previsibilidade e capacidade de adaptação frente ao contexto em que se dá o trabalho.

Com o amadurecimento das metodologias de desenvolvimento de software, surgiram também ferramentas de apoio a sua implementação, auxiliando no rastreamento e documentação do processo de desenvolvimento. Entre as informações registradas por tais ferramentas podemos encontrar descrição de requisitos, métricas associadas à realização do trabalho e interações entre pessoas envolvidas. Jira¹ é uma ferramenta auxiliar para gerenciamento de projetos de software com grande adoção por parte de empresas e comunidades open-source.

Sendo o desenvolvimento de software uma atividade que exige considerável esforço cognitivo por parte das pessoas envolvidas, fatores como o estado emocional dos indivíduos pode ter impacto na produtividade do time e na qualidade das soluções desenvolvidas (GRAZIOTIN et al., 2017). Buscando mensurar a extensão deste impacto, diversos estudos de análise de sentimento em artefatos de software foram conduzidos. Por exemplo, Ortu et al. (2015) analisou o sentimento expresso em uma base de dados de cartões em projetos de código aberto, e descobriu relação entre o sentimento expresso na descrição de uma tarefa e seu tempo de resolução. Já em outro estudo Mäntylä et al. (2016), um corpus similar é analisado com finalidade de identificar sinais de estresse e perda de produtividade.

Bay Bridge é uma grande empresa varejista do setor de moda que mantém sua própria divisão de tecnologia. A divisão de tecnologia é responsável pelo desenvolvimento de sistemas abertos ao público consumidor, bem como sistemas de uso interno que dão suporte à operação da empresa. A divisão de tecnologia adota o framework Scrum (SCRUM, 2018) para gerenciamento de trabalho, desenvolvendo seus produtos de software de maneira iterativa e incremental. Jira é utilizada com ferramenta de auxílio no gerenciamento de projetos.

¹<https://www.atlassian.com/software/jira>

Ao longo de um ciclo de desenvolvimento, tarefas são registradas no Jira na forma de cartões, que representam requisitos de novas funcionalidades, registro de defeitos, dúvidas técnicas, etc. Cada cartão contém informações como título, descrição detalhada, status atual do trabalho, etc. Além disso, cartões Jira têm uma seção reservada a comentários, que serve como meio para discussões relacionadas ao cartão.

O objetivo do presente trabalho é verificar a existência de correlação entre sentimento associado a um cartão, na forma de seu enunciado e comentários, e métricas de produtividade e qualidade. Além disso busca-se verificar se características da tarefa relacionam-se com o sentimento expresso no cartão. Para tanto foram utilizados dados de um projeto extraídos da ferramenta Jira utilizada pela organização Bay Bridge.

De maneira mais objetiva, este trabalho busca responder às seguintes perguntas dentro do contexto da organização Bay Bridge:

- A assertividade na descrição de um cartão influencia o tempo necessário para sua resolução?
- A assertividade na descrição de um cartão influencia o número de rejeições na etapa de validação?
- Existe relação entre assertividade na descrição de um cartão e o sentimento expresso em seus comentários?
- Existe variação de sentimento em função do tipo de trabalho (nova funcionalidade vs. defeito)?

Diferentemente dos trabalhos já realizados nesta área, que usam projetos open-source (e.g. (ORTU et al., 2015; MäNTYLÄ et al., 2016), o presente trabalho analisa sentimento em artefatos de software produzidos dentro de um contexto corporativo.

O restante deste trabalho é organizado como segue:

- No Capítulo 2 são apresentados fundamentos teóricos relacionados a análise de sentimento, framework Scrum e a ferramenta de rastreamento de projetos Jira.
- No Capítulo 3 são apresentados o contexto da organização analisada e como se dá a utilização do framework Scrum e Jira.
- No Capítulo 4 é apresentada a metodologia empregada na análise de sentimento.
- No Capítulo 5 são apresentados e discutidos os resultados obtidos na análise.
- No Capítulo 6 são apresentadas considerações finais sobre o trabalho.

2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo trata de conceitos relacionados à análise de sentimento e trabalhos relacionados à sua aplicação na análise de artefatos de software.

2.1 Análise de Sentimento

Análise de sentimento diz respeito a qualquer estudo computacional que tem como alvo a subjetividade expressa em forma textual (LIU, 2010). Entende-se por subjetividade manifestações de afeição, emoção, opinião, posicionamento, entre outras. Dado que existem diversas manifestações de subjetividade, a área de análise de sentimento compreende também diferentes abordagens para medi-la.

Uma abordagem popular para análise de sentimentos é o estudo da polaridade, que busca mensurar o nível de afeição ou interesse do autor em relação ao tópico sobre o qual escreve. Um possível resultado de análise de polaridade seria a atribuição de categorias de polaridade positiva, negativa ou neutra a um determinado trecho de texto.

Outra abordagem comumente utilizada é o estudo de emoções, que tem por objetivo determinar qual tipo de emoção está presente no texto analisado segundo uma teoria de emoções. Como exemplo, poderíamos atribuir rótulos de emoção (alegria, tristeza, raiva, etc) a um determinado trecho de texto de acordo com o modelo de emoções.

Neste trabalho, escolhemos estudar sentimento na forma de emoção.

2.2 Frameworks para Teorias de Emoção

A análise de sentimento na forma de estudo de emoções consiste em caracterizar a emoção contida em um texto de acordo com uma teoria de emoções. Segundo Graziotin, Wang and Abrahamsson (2015), teorias existentes para estudo de emoções podem ser agrupadas em dois tipos de frameworks conceituais que compartilham grande similaridade em suas abordagens, a saber, discreto e dimensional, detalhados abaixo.

2.2.1 Framework Discreto

A abordagem discreta tem por principal característica enumerar um conjunto de emoções básicas que podem ser percebidas individualmente. Como exemplo de teoria pertencente ao framework discreto, temos o modelo proposto por Ekman (GRAZIOTIN et al., 2017), que apresenta seis emoções básicas: aversão, felicidade, medo, raiva, surpresa e tristeza.

Outra teoria discreta amplamente referenciada é o modelo circular de emoções de Plutchik (MOHAMMAD; TURNEY, 2013) que, além de estender o modelo de Ekman com emoções básicas de confiança e expectativa, inclui noções de variação de intensidade e combinação de emoções básicas como forma de originar emoções secundárias.

2.2.2 Framework Dimensional

A abordagem dimensional tem por principal característica utilizar dimensões ortogonais na caracterização de emoções, as quais são mensuradas em uma escala contínua.

O modelo mais popular desta abordagem é o VAD, descrito inicialmente em (MEHRABIAN, 1996), no qual o estado emocional de um indivíduo pode ser descrito através de três dimensões: valência (valence), excitação (arousal) e dominância (dominance). A dimensão de valência representa o quão agradável ou desagradável é o estímulo que originou a emoção. Excitação representa a intensidade da resposta emocional ao estímulo. Dominância representa a posição percebida em relação ao estímulo no que diz respeito a controle ou submissão. Uma emoção em particular é representada pela atribuição de valores para cada uma das dimensões, em uma escala que vai de 0 a 10.

Neste trabalho, optamos por utilizar o modelo de emoções VAD por conta da disponibilidade de dicionários de emoção de propósito geral.

2.3 Técnicas de Análise de Sentimento

As técnicas mais aplicadas na análise de sentimento segundo Tsytsarau and Palpanas (2012) são a baseada em aprendizado de máquina supervisionado e a léxica.

A técnica baseada em aprendizado de máquina supervisionado consiste em treinar um algoritmo de classificação usando um corpus de textos anotados, ou seja, um conjunto

de textos onde a cada um é atribuído um rótulo representando o sentimento presente no mesmo. Nesta abordagem, o algoritmo é busca identificar padrões implícitos presentes no conjunto de dados de treinamento e estabelecer relação com os respectivos rótulos. Ao ser apresentado a um novo texto, o modelo de classificação resultante é analisa o texto e classifica-o de acordo com as relações estabelecidas durante a fase de treinamento. Esta abordagem requer um corpus de textos anotados específicos para o domínio, já que os resultados obtidos são sensíveis em relação à quantidade e qualidade dos dados usados para treinamento.

A abordagem baseada em dicionários, também conhecida como abordagem léxica, utiliza-se de um dicionário de sentimentos, isto é, uma compilação de termos associados a suas respectivas avaliações de sentimento. No trabalho de Kolchyna et al. (2015), são citadas três abordagens para a construção de léxicos de sentimento. A abordagem direta para construção é através da atribuição manual de avaliações de sentimento a cada termo, feita por seres humanos. Esta abordagem tem por vantagem ser de simples implementação, mas geralmente demanda considerável investimento de tempo. A segunda abordagem consiste da construção automática do léxico a partir de um corpus de textos anotados utilizando técnicas de aprendizado de máquina supervisionado. Esta abordagem tem a vantagem de ser totalmente automatizada, mas requer um corpus de texto anotado. A última abordagem consiste na expansão de um léxico existente através da exploração das relações existente entre palavras. O léxico SentiWordNet (BACCIANELLA; ESULI; SEBASTIANI, 2010) foi construído a partir de um conjunto de termos semente e expandido através das relações extraídas do recurso de WordNet (FELLBAUM, 1998).

Na abordagem de análise de sentimentos baseada em dicionários, além de atribuir um sentimento a termos/expressões, é preciso definir também uma função de agregação, cuja responsabilidade é compor a avaliação de sentimento de um texto completo a partir da avaliação dos termos individuais presentes.

Existem ferramentas para análise de sentimentos de propósito geral disponíveis no mercado. Entre elas destaca-se a ferramenta para análise de polaridade SentiStrenght (THELWALL et al., 2010), que possui seu próprio léxico e função para cálculo de polaridade agregada. A ferramenta divide o texto em segmentos e, a cada segmento, atribui uma avaliação de polaridade baseada nos valores máximos e mínimos de polaridade dos termos presentes no segmento.

Neste trabalho optamos por utilizar a abordagem baseada em dicionário, pela ausência de um corpus anotado de tamanho adequado no domínio de desenvolvimento de

software.

2.4 Dicionários de Emoção

Um dicionário de sentimentos associa a cada termo um sentimento de acordo com o framework adotado. Para emoções, temos como exemplo os dicionários NRC e o ANEW. Todos estes dicionários são voltados à língua inglesa, embora existam esforços de tradução do NRC.

O dicionário NRC (MOHAMMAD; TURNEY, 2013) é um léxico baseado no modelo de oito emoções de Plutchik. Para cada um dos termos presentes, o léxico lista a presença ou ausência de cada emoção básica associada ao termo. Além disso, o dicionário também apresenta a avaliação de polaridade do termo. O léxico NRC contém aproximadamente dez mil termos em Inglês e foi construído utilizando o serviço Mechanical Turk da Amazon, uma plataforma que permite a resolução de tarefas que necessitam inteligência humana.

O dicionário ANEW (BRADLEY; LANG, 1999) é um léxico de emoções que contém avaliações para as dimensões de valência, excitação e dominância para cerca de mil palavras. Em trabalho posterior de Warriner, Kuperman and Brysbaert (2013), este léxico foi expandido para mais de 13 mil termos, incluindo termos associados a categorias específicas (doenças, profissões, palavras tabu, etc). Além de expandir o número de categorias contempladas e números de termos presentes, o léxico agrega as avaliações de emoção por idade, gênero e nível de escolaridade. Na tabela 2.4, temos alguns exemplos de avaliações de emoção extraídas da versão estendida do léxico ANEW.

Tabela 2.1: Exemplos de avaliação VAD

Termo	Valência	Excitação	Dominância
eager	6.37	5.78	6.46
upset	2.45	4.49	4.30
house	7.19	3.95	6.41
school	5.41	4.57	5.88
sing	7.50	4.10	6.67
try	5.64	4.00	5.94

Neste trabalho, por termos escolhido o framework VAD, utilizaremos a versão expandida do dicionário ANEW.

2.5 Processamento de Linguagem Natural

Análise de sentimentos tem por objeto de estudo textos escritos em linguagem humana. Assim é necessária a aplicação de técnicas de processamento de linguagem natural (PLN). O PLN trata da avaliação computacional de diversos aspectos da comunicação humana, tais como estrutura, significado, e contexto (BECKER, 2014). As principais técnicas de PLN aplicada neste trabalho foram tokenização e lematização.

Tokenização é a técnica de identificar termos (tokens) que compõem um texto. Embora abordagens para extração de tokens possam variar de acordo com a linguagem e objetivo da aplicação, a ideia central é de isolar os termos através da remoção de caracteres separadores (espaços, tabulações, quebra de linha, pontuações, etc).

Lematização consiste em agrupar diferentes variações de um termo sob uma representação canônica chamada lema. Assim palavras pertencentes ao mesmo grupo podem ser avaliadas de maneira única através do uso do lema. Como exemplo, temos que "cantou", "cantarei" e "cantamos" são agrupadas sob o lema "cantar".

2.6 Scrum

Scrum é um framework de gerenciamento de trabalho para desenvolvimento de produtos, que adota uma abordagem iterativa e incremental (RUBIN, 2012). Ao evitar planejamento com grande antecipação, o framework permite flexibilidade e capacidade de adaptação às demandas de negócio. O framework Scrum é caracterizado por um conjunto de papéis, artefatos e cerimônias dispostas ao longo de um ciclo de desenvolvimento.

2.6.1 Papéis

Cada membro de um time Scrum desempenha um papel com atribuições específicas, são eles: product owner, scrum master e desenvolvedor. O product owner é responsável por incorporar a perspectiva do negócio no processo de desenvolvimento, garantindo que o time foca esforços na entrega de valor real. Scrum master garante que o framework Scrum é aplicado corretamente, reforçando os conceitos junto ao time e facilitando cerimônias. Além disso o scrum master também age como uma camada de proteção contra fatores externos e trabalha na remoção de impedimentos. Desenvolvedores são responsá-

veis pela execução e entrega de tarefas técnicas. O número de desenvolvedores em um time scrum varia usualmente de três a nove desenvolvedores.

2.6.2 Artefatos

O framework Scrum prevê três tipos de artefatos: o backlog de produto, o backlog da sprint e o incremento. O backlog do produto é conjunto de requisitos do produto, normalmente composto por histórias de usuário, reporte de defeitos e requisitos não funcionais. O backlog do produto é gerenciado pelo product owner que decide quais itens devem ser incluídos e a prioridade relativa entre eles. Já o backlog da sprint é o conjunto de requisitos que devem ser implementados pelo time durante o ciclo de desenvolvimento. Incremento de produto é a soma de todos os itens do backlog de sprint resolvidos durante um ciclo de desenvolvimento.

2.6.3 Ciclo de Desenvolvimento Scrum

Os itens do backlog da sprint devem ser implementados durante um ciclo de desenvolvimento de tamanho fixo chamado sprint. Ao longo da sprint, além das tarefas de desenvolvimento, devem ser realizadas cerimônias com propósitos específicos.

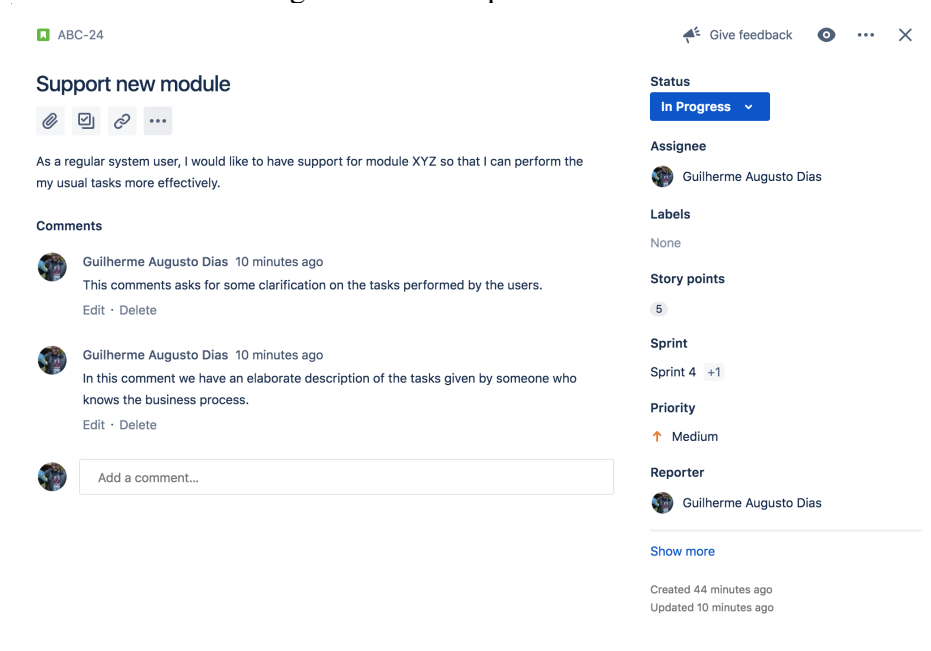
A reunião de planejamento de sprint acontece no início da sprint e tem por principal objetivo selecionar os itens do backlog de produto que devem ser incluídos no backlog da sprint. A seleção dos itens é feita considerando prioridades de negócio e capacidade de entrega do time. Standup é uma reunião diária onde cada membro do time de desenvolvimento informa o que foi feito no dia anterior, o planejamento para o dia atual e se existe algum impedimento. A visibilidade proveniente das reuniões de standup permite o re-planejamento durante a execução do trabalho. No fechamento da sprint, acontece a reunião de revisão de sprint onde o time e stakeholders do produto revisam o incremento de produto entregue e discutem objetivos de negócio para o próximo sprint. Após a revisão de sprint e antes da próxima reunião de planejamento de sprint, acontece a retrospectiva, onde o time discute o que correu bem, quais são os pontos de melhoria e comprometem-se a trabalhar em um sub-conjunto dos pontos de melhoria durante o próximo ciclo.

2.7 Jira

Jira é uma ferramenta de auxílio para gerenciamento de projetos desenvolvida pela Atlassian. Jira é uma ferramenta de propósito amplo, que permite a implementação de diversos métodos de trabalho através de seus módulos, sendo apenas as funcionalidades relacionadas ao módulo Scrum de interesse deste trabalho.

O módulo de suporte a Scrum do Jira facilita o gerenciamento dos artefatos previstos pelo framework Scrum, bem como permite acompanhar progresso em tarefas dentro de uma sprint segundo um fluxo de trabalho. Os artefatos Scrum gerenciados pelo Jira são os itens do backlog de produto e backlog da sprint, representados na forma de cartões. A cada cartão podemos associar atributos variados, por exemplo, tipo de trabalho (funcionalidade vs. defeito), prioridade, estimativa de esforço, etc. A Figura 2.1 ilustra um cartão hipotético.

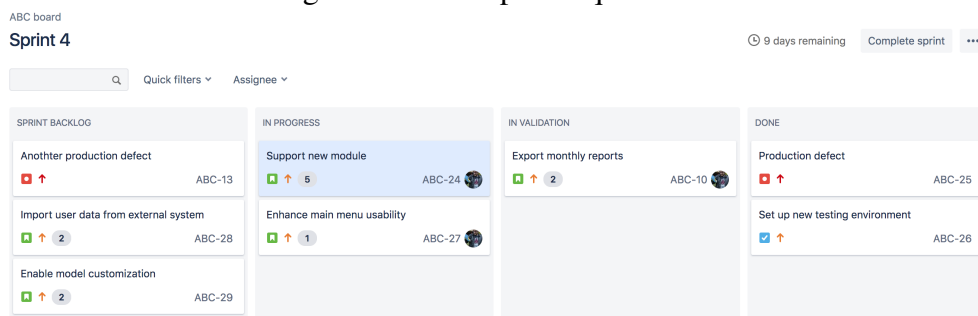
Figura 2.1: Exemplo de cartão Jira



Além dos campos que caracterizam a tarefa representada pelo cartão, existe também uma seção reservada a comentários, onde pessoas envolvidas na resolução do cartão podem comunicar-se.

Uma vez incluído no backlog da sprint, é possível acompanhar o trabalho em andamento para cada cartão através das alterações de status em um workflow pré-definido. O acompanhamento se dá através de um quadro onde é possível visualizar todos os cartões incluídos na sprint corrente dispostos em colunas correspondentes a seu status atual, tal como ilustrado na Figura 2.2.

Figura 2.2: Exemplo de quadro Jira



2.8 Trabalhos Relacionados

Nesta seção apresentamos trabalhos relacionados à análise de sentimentos aplicada a artefatos de software. Os trabalhos escolhidos visam associar métricas associadas ao processo de desenvolvimento de software ao sentimento existentes em ferramentas auxiliares para gerenciamento de projetos.

Mäntylä et al. (2016) propõem a análise de emoções presentes em artefatos de software com o objetivo de identificar estresse. O trabalho é baseado no conjunto de dados que consiste de aproximadamente 700 mil cartões Jira provenientes de projetos open-source. O estudo segue a abordagem baseada em dicionários, utilizando a versão expandida do léxico ANEW. A função de agregação utilizada é uma adaptação da fórmula de avaliação de polaridade implementada pelo SentiStrength, atribuindo a um texto a avaliação de emoção correspondente à diferença entre os valores extremos encontrados no mesmo. Os resultados deste trabalho sugerem que é possível detectar variação emocional das pessoas envolvidas ao longo do ciclo de vida de um cartão. O estudo observa que períodos prolongados de pressão, materializados na forma de cartões com alta prioridade e longo tempo de resolução, estão associados a altas medidas de excitação e baixas medidas de valência por parte da pessoa trabalhando na resolução do cartão. A combinação de tais valores para dimensões de excitação e valência é tida no estudo como indicadora de um estado de estresse. Também foi possível encontrar correlação entre determinadas características do cartão, tais como tipo de trabalho (feature vs. defeito) e prioridade, e a emoção expressa no conteúdo do cartão. Além disso, é proposta a hipótese de que valores maiores de dominância teria impacto no tempo de resolução de cartões, com maior dominância implicando um tempo menor de resolução. Os resultados do estudo, no entanto, não mostram relação significativa entre dominância e tempo de resolução.

Já Ortu et al. (2015) buscam estabelecer relação entre sentimentos associados a

uma tarefa e seu tempo de resolução. Para isso, analisam um conjunto de dados que consiste de aproximadamente 560 mil cartões Jira provenientes de projetos open-source. A técnica de análise de sentimentos empregada combina informações de polaridade, cortesia e emoções. A informação de polaridade é extraída com auxílio da ferramenta SentiStrength. Para extração de informação de cortesia foi utilizado a ferramenta baseada em aprendizado de máquina desenvolvida por Danescu-Niculescu-Mizil et al. (2013). A análise de emoções é feita através de um classificador desenvolvido sobre um corpus que foi anotado para este fim. O modelo atribui rótulos de emoção segundo um modelo de emoções discreto contendo quatro emoções: alegria, amor, raiva e tristeza. Os resultados mostram que cartões contendo emoções positivas (alegria e amor) têm tempo de resolução menor que cartões contendo tristeza. Além disso, foi possível observar que cartões com manifestações extremas de cortesia ou descortesia têm tempo de resolução significativamente menor que cartões que não apresentam tais sentimentos.

3 CONTEXTO DA ORGANIZAÇÃO BAY BRIDGE

3.1 Organização

A organização em questão é uma empresa varejista do setor de moda com sede nos Estados Unidos, mas atuando em diferentes mercados ao redor do globo. Buscando maximizar sua parcela de mercado em cada localidade onde atua, a empresa implementa algumas ações estratégicas. Entre elas vale destacar:

- Múltiplas marcas: sabendo que o público consumidor de moda é composto por diversos nichos, cada um com suas particularidades, a empresa apresenta seus produtos sob diferentes marcas. Cada marca tem sua própria identidade e busca atender as necessidades de um público específico.
- Múltiplos canais: com a consolidação do consumo através de canais digitais, a empresa busca engajamento com consumidores através destes também, por exemplo, com seu próprio sistema de e-commerce.

Sendo uma empresa que opera em escala global, gerenciando uma estrutura organizacional complexa e buscando engajamento com consumidores através de canais digitais, o uso e desenvolvimento de produtos de software torna-se imprescindível para o sucesso da empresa. Para atender tais demandas foi criada uma divisão interna de tecnologia.

A divisão de tecnologia trabalha em parceria com os demais setores da empresa para prover soluções para o negócio. Entre os serviços prestados ao restante da organização temos desenvolvimento de produtos de software, gerenciamento de infraestrutura de tecnologia e suporte técnico. Como o escopo deste trabalho limita-se a análise dos processos de desenvolvimento de software dentro da organização, aspectos relacionados a gerenciamento de infraestrutura e suporte técnico não serão abordados.

A subdivisão de desenvolvimento organiza-se em linhas de produto. Uma linha de produto é um agrupamento lógico de produtos de software que colaboram para atender uma necessidade específica do negócio. Como exemplos temos as linhas de produto de planejamento e vendas.

Para cada produto de software existe um time de desenvolvimento responsável pela manutenção do sistema e implementação de novas funcionalidades. Os times executam seu trabalho de maneira independente, definindo quais tecnologias utilizar, a arquite-

tura da solução e demais aspectos.

Apesar de atuarem com considerável autonomia, os times de desenvolvimento devem seguir algumas diretrizes da divisão de tecnologia. Atualmente existe uma norma relacionada a processo de software que estabelece uma variante do Scrum como o padrão dentro da subdivisão de desenvolvimento. Além de estabelecer os padrões relacionados ao processo de desenvolvimento, a divisão de tecnologia também regula o uso de ferramentas auxiliares ao processo, sendo Jira o sistema de rastreamento de tarefas que deve ser adotado por todos os times de produto.

3.2 Processo de Desenvolvimento

O processo padrão definido pela divisão de tecnologia para desenvolvimento de software é o Scrum. Segue a descrição de como cada um dos aspectos do Scrum relevantes a este trabalho é implementado pela subdivisão de desenvolvimento.

3.2.1 Backlog do produto

O backlog do produto é construído durante uma atividade conhecida como concepção. A concepção é um conjunto de exercícios de descoberta que tem como principais objetivos definir funcionalidades essenciais para o produto e a prioridade relativa entre elas. Participam da concepção o scrum master, o gerente de produto, o time de desenvolvimento e representantes do negócio. O produto final de uma atividade de concepção é um conjunto de épicos representando uma funcionalidade e as tarefas que constituem os épicos. As tarefas são então registradas como cartões no Jira e passam a integrar o backlog do produto. Neste momento os cartões contêm não mais que um título e uma breve descrição da tarefa, visto que informações mais detalhadas serão registradas em etapas posteriores do processo, principalmente na reunião de planejamento de sprint.

Além de cartões provenientes da atividade de concepção, novos cartões podem ser adicionados ao backlog de produto a qualquer momento. Este tipo de adição compreende majoritariamente trabalho associado ao conserto de defeitos, mas podem também representar tarefas de melhoria técnica ou mudanças de requisitos.

Cada vez que finaliza-se a implementação das funcionalidades da concepção, é organizada uma nova atividade de concepção para definir próximas funcionalidades a se-

rem trabalhadas. A duração do ciclo que se inicia com a realização de uma concepção e termina com a entrega de todas as funcionalidades previstas na mesma, dura aproximadamente 3 meses.

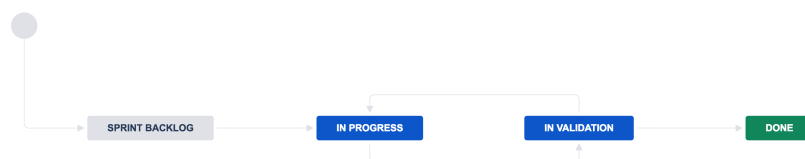
3.2.2 Ciclos de desenvolvimento

O desenvolvimento das funcionalidades priorizadas durante a atividade de concepção se dá em ciclos de duas semanas chamados sprints. Ao longo de uma sprint, além das tarefas de desenvolvimento, são realizadas as cerimônias de planejamento de sprint, revisão de sprint e retrospectiva.

A sessão de planejamento de sprint é onde se determinam o objetivo para a próxima sprint e quais tarefas de desenvolvimento devem ser completadas para alcançá-lo. Participam da sessão de planejamento de sprint o scrum master, o gerente de produto e o time de desenvolvimento. Durante esta sessão, o gerente de produto expõe as demandas de maior prioridade com base em feedback recebido de representantes de negócio e usuários do produto. É então selecionado, a partir do backlog de produto, um conjunto de cartões que atendam as necessidades do negócio, podendo estes ser novas funcionalidades ou conserto de defeitos. Aos cartões selecionadas são atribuídas estimativas de esforço na forma de pontos que variam de 0 (esforço mínimo) a 8 (esforço grande suficiente para subdividir a tarefa), bem como uma avaliação de prioridade, variando entre prioridade baixa, média e alta. Uma vez definidos a prioridade e estimativa de esforço para cada cartão selecionado, estes valores são registrados no Jira.

O progresso de trabalho em cartões durante uma sprint é acompanhado através das mudanças de status ao longo de um fluxo de trabalho pré-determinado. O fluxo de trabalho consiste dos possíveis status que podem ser atribuídos a um cartão e as regras de transição entre os eles.

Figura 3.1: Fluxo de trabalho da Bay Bridge



A Figura 3.1 ilustra o fluxo de trabalho adotado pela Bay Bridge. A cartões recentemente incluídos na sprint são atribuídos o status *Sprint Backlog*. Ao iniciar o desenvolvimento da tarefa, o desenvolvedor deve atualizar o status no Jira para *Em Progresso*.

Uma vez finalizado o desenvolvimento da tarefa, o desenvolvedor responsável, atualiza o status para *Em Validação* e um testador designado será responsável por validar a solução desenvolvida. Caso a solução seja satisfatória, o testador atualiza o status para *Finalizado*, resolvendo efetivamente o cartão. No caso em que falhas de implementação sejam descobertas durante a etapa de validação, o testador retorna o cartão para o status *Em Progresso* para que o desenvolvedor responsável possa sanar o problema.

3.3 Motivação

O time Horse é responsável pelo desenvolvimento de um dos produtos da organização Bay Bridge. Neste time, os cartões são escritos tanto por gerentes de produto quanto por desenvolvedores, gerando uma grande diversidade no padrão de escrita dos cartões. Tal diversidade pode ser considerada uma espaço para melhoria, uma vez que um padrão de qualidade para descrição de cartões Jira poderia ter um impacto positivo na produtividade do time, qualidade da solução e gerar uma menor carga cognitiva no momento de compreender a tarefa a ser executada.

O componente de dominância em uma emoção no modelo VAD representa a sensação de controle em relação ao tópico sobre que se discorre. A hipótese elaborada neste trabalho é de que quanto maior o conhecimento da pessoa responsável pela criação do cartão sobre o domínio do problema, maior será a medida de dominância expressa no texto do cartão. Ao mesmo tempo, a pessoa com maior conhecimento do domínio é capaz de produzir um cartão com descrição mais assertiva, facilitando o posterior trabalho no desenvolvimento do cartão. Tal hipótese é também formulada e verificada no trabalho desenvolvido por Mäntylä et al. (2016). Assim a motivação deste trabalho é verificar se a assertividade na descrição do cartão, medida através da dominância contida no texto, tem impacto na produtividade do time e qualidade da solução, bem como no sentimento das pessoas envolvidas na resolução da tarefa.

4 MÉTODO

Neste Capítulo são apresentadas as análises propostas e suas motivações, bem como os métodos de coleta de dados, tratamentos aplicados e análise de sentimento utilizados neste trabalho.

4.1 Análises propostas

As análises propostas neste trabalho podem ser divididas em três categorias. A primeira categoria refere-se à relação entre o sentimento expresso no cartão e métricas de produtividade e qualidade. A motivação por trás desta análise é a hipótese de que cartões com descrições precisas e assertivas seriam mais facilmente resolvidas e teriam menores taxas de rejeição em etapas de validação. Mais especificamente queremos verificar como a assertividade do conteúdo de um cartão, expressa através da dimensão de dominância do modelo VAD conforme explicado na Seção 3.3, impacta o tempo de resolução e o número de rejeições na etapa de validação. Esta primeira categoria de análise está relacionada às duas primeiras perguntas enunciadas no Capítulo 1:

- A assertividade na descrição de um cartão influencia o tempo necessário para sua resolução?
- A assertividade na descrição de um cartão influencia o número de rejeições na etapa de validação?

Para responder a primeira pergunta, verificou-se a correlação entre a medida de dominância no título e a descrição do cartão com o tempo necessário para resolução do mesmo. Para responder a segunda pergunta, foi feita uma comparação entre as médias da dominância contidas no título e descrição de cartões que não foram rejeitados e cartões que foram rejeitados uma ou mais vezes na etapa de validação.

A segunda categoria diz respeito à relação entre a assertividade contida na descrição do cartão e o sentimento das pessoas envolvidas na resolução do cartão. A hipótese que gostaríamos de verificar é que cartões redigidos de maneira assertiva teriam impacto na emoção de pessoas envolvidas na resolução da tarefa, implicando na observação de correlação entre as avaliações de emoção da descrição dos comentários. Esta categoria de análise diz respeito à terceira pergunta enunciada no Capítulo 1:

- Existe relação entre assertividade na descrição de um cartão e o sentimento expresso

em seus comentários?

Para responder a esta pergunta, foi verificada a correlação entre a medida de dominância na descrição do cartão e o sentimento representativo dos comentários do cartão. O sentimento representativo consiste dos maiores valores para cada uma das dimensões do modelo VAD encontrada nos comentários do cartão.

A última categoria diz respeito à relação entre a natureza do trabalho representado pelo cartão e o sentimento expresso no mesmo. Este grupo de análise parte do pressuposto de que características intrínsecas do trabalho a ser realizado podem ter impacto na maneira como os desenvolvedores se sentem, e que tal sentimento é expresso nas comunicações que se dão através da ferramenta Jira. Este grupo de análise está relacionado a última pergunta apresentada no Capítulo 1:

- Existe variação de sentimento em função do tipo de trabalho (nova funcionalidade vs. defeito)?

Para verificar relação entre sentimento e tipo de trabalho foi feita uma comparação entre as médias de avaliação de sentimento, para cada uma das dimensões do modelo VAD, entre cartões representando novas funcionalidades e defeitos.

4.2 Coleta de Dados

A base de dados analisada é composta por informação extraída a partir da instância Jira utilizada pela divisão de tecnologia da organização Bay Bridge. Os dados coletados compreendem o histórico completo de cartões associados ao projeto Horse. Mais especificamente, os dados correspondem a cartões Jira do projeto Horse criados no período de 20 de Fevereiro de 2017 a 5 de Dezembro de 2017, totalizando 4018 cartões.

A extração dos dados foi automatizada através do uso da interface de programação de aplicações expostas pelo serviço Jira. A informação recuperada para cada cartão inclui:

- Título: uma descrição resumida do trabalho contido no cartão.
- Descrição: uma descrição detalhada do trabalho contido no cartão.
- Status: representa o estágio de desenvolvimento em que se encontra o cartão, podendo assumir os valores ilustrados na figura 3.1.
- Tipo: contém o tipo de tarefa representada pelo cartão, podendo variar entre épico, nova funcionalidade, defeito, investigação e dívida técnica.

- **Prioridade:** representa a prioridade relativa entre cartões dentro de uma sprint. Assume valores inteiro de 1 a 3, sendo 1 a mais alta prioridade e 3 a mais baixa prioridade.
- **Pontuação:** os pontos atribuídos a um cartão representam a estimativa de esforço necessário para concluir a tarefa nele contida. A escala de pontuação permite valores de 0, 1, 2 e 5 pontos, onde 0 representa uma tarefa trivial e 5 o nível de complexidade mais elevado.
- **Comentários:** a lista de comentários associados ao cartão.
- **Histórico de alterações:** a lista de todas as alterações realizadas sobre o cartão. Cada alteração consiste de um identificador do atributo sendo alterado, seu valor original, o novo valor e a data em que a modificação foi realizada.

4.3 Cálculo de Sentimento

A análise de sentimento neste trabalho consiste de uma medida de emoção baseada no modelo de emoções VAD (Valence-Arousal-Dominance), pertencente a classe de teorias dimensionais. Para medir a emoção contida nos textos dos cartões, foi adotada uma abordagem baseada em dicionários. O cálculo de emoção para um texto consiste de duas etapas descritas abaixo.

Na primeira fase do cálculo de emoção, são recuperadas as avaliações VAD para cada um dos termos contidos no texto. A informação de emoção é recuperada a partir de uma consulta a um dicionário de emoções. O léxico de emoções utilizado foi a versão estendida do ANEW, descrita na Seção 2.4. Termos cuja avaliação de emoção não consta no léxico de emoções utilizado são desconsiderados na análise.

A segunda etapa do cálculo de emoção consiste em combinar a avaliação VAD de cada um dos termos do termo considerados para a análise através do uso de uma função de agregação. A função de agregação recebe como entrada a coleção de avaliações de emoção dos termos individuais considerados e produz um valor representativo para a avaliação de emoção do texto completo. Critérios utilizados para agregar avaliações de emoção individuais podem ser variados e devem atender à necessidade da análise pretendida. Neste trabalho foram utilizadas duas funções de agregação distintas que são apresentadas no restante desta Seção.

4.3.1 Amplitude máxima

A função de agregação por amplitude máxima usa a variação de emoção entre o conjunto de palavras utilizadas no texto como representante da quantidade de emoção associada ao texto. Esta estratégia de agregação não preserva informação sobre a orientação do sentimento. Para a dimensão de valência, ela indica a quantidade de emoção que o texto apresenta nesta dimensão, mas não provê informação nenhuma a respeito da positividade ou negatividade desta emoção. O mesmo para as dimensões dominância e excitação, onde ela não indica quão alta ou baixa é a intensidade da emoção nestas emoções. Usaremos a agregação por amplitude máxima para verificar a relação entre tipo de trabalho e emoção contida nos cartões, já que para esta análise nos interessa apenas saber se existe diferença na intensidade das ativações emocionais associadas a cada tipo de trabalho e não necessariamente sua orientação.

Considerando uma dimensão do modelo VAD em isolamento, temos que a função de agregação baseada na amplitude máxima produz a diferença entre o maior e o menor valor de avaliação de emoção presentes no texto analisado. Em casos onde a avaliação de emoção máxima presente no texto é menor que a média das avaliações de todos os termos contidos no léxico, assumimos que o valor máximo equivale a média do léxico. De maneira análoga, quando a avaliação de emoção mínima presente no texto é maior que a média das avaliações de todos os termos contidos no léxico, assumimos que o valor mínimo equivale à média do léxico.

Seja \bar{w} a lista de avaliações de emoção para os termos do texto analisado e \bar{W} a lista de avaliações de emoção para os termos contido no léxico. A função de agregação por amplitude máxima para uma dada dimensão do modelo VAD pode ser definida pela seguinte equação:

$$\max Amp(\bar{w}) = \begin{cases} \max(\bar{w}) - \text{avg}(\bar{W}), & \text{se } \min(\bar{w}) > \text{avg}(\bar{W}) \\ \text{avg}(\bar{W}) - \min(\bar{w}), & \text{se } \max(\bar{w}) < \text{avg}(\bar{W}) \\ \max(\bar{w}) - \min(\bar{w}), & \text{se } \min(\bar{w}) \leq \text{avg}(\bar{W}) \leq \max(\bar{w}) \end{cases} \quad (4.1)$$

4.3.2 Maior distância da média

Diferentemente da função que emprega amplitude máxima, a função de agregação por maior distância da média preserva informação do posicionamento da emoção em relação a um valor âncora (a média de avaliações do léxico), permitindo diferenciar os casos com avaliações baixas, dos casos com avaliações altas. Tomando novamente a dimensão de valência como exemplo, temos que a agregação por maior distância da média produz valores negativos para textos com carga emocional negativa e valores positivos para textos com carga emocional positiva. Nas demais dimensões, ele permite distinguir casos abaixo da média, ou acima da média. Usaremos a agregação por maior distância da média nos cenários que envolvem a correlação de sentimento com métricas do cartão e com o sentimento dos comentários, uma vez que é preciso diferenciar avaliações emocionais baixas de avaliações emocionais altas para poder verificar tal correlação.

Para uma dada dimensão do modelo VAD, a função de agregação baseada em maior distância da média, calcula para cada avaliação de emoção presente no texto analisado sua distância em relação à média das avaliações de todos os termos contidos no léxico. A distância é o valor absoluto da diferença entre a avaliação e a média do léxico. O valor agregado de emoção corresponde a diferença entre a avaliação de emoção mais distante da média e a própria média.

Seja \bar{w} a lista de avaliações de emoção para os termos do texto analisado, \bar{W} a lista de avaliações de emoção para os termos contido no léxico e $maxMeanDist$ uma função auxiliar que, dado uma lista de avaliações de emoção, retorna a avaliação com maior distância da média. A função de agregação por maior distância da média para uma dada dimensão do modelo VAD pode ser definida pela equação:

$$meanDist = maxMeanDist(\bar{w}) - avg(\bar{W}) \quad (4.2)$$

4.4 Tratamento aos Dados

De modo a facilitar a análise, os dados inicialmente coletados foram submetidos a uma série de tratamentos. Os tratamentos consistem basicamente de derivações de informações presentes no histórico de alterações do cartão. As informações derivadas desta maneira foram: o tempo de resolução de um cartão e o número de rejeições em etapa de validação.

4.4.1 Tempo de resolução

O tempo de resolução de um cartão é definido neste trabalho como o tempo transcorrido do momento em que um cartão tem seu status alterado para *Em progresso* pela primeira vez até o momento quando é movido para o status *Finalizado*.

O histórico de alterações do cartão registra mudanças de status, juntamente com a data e hora em que se deu a alteração, sendo assim possível, determinar o tempo transcorrido durante as mudanças de status de interesse. Na figura 4.1 apresenta um exemplo de alteração de status no histórico de um cartão.

Figura 4.1: Entrada no histórico de alterações

▼ [25]	{ 4 fields }
id	6684960
▶ author	{ 8 fields }
created	2017-09-25T10:13:43.174-0700
▼ items	[1 element]
▼ [0]	{ 6 fields }
field	status
fieldtype	jira
from	10017
fromString	Ready to Play
to	3
toString	In Progress

O intervalo de tempo consiste do número de horas transcorridos entre as alterações de status de interesse, considerando um dia de trabalho de oito horas e excluindo finais de semana.

4.4.2 Número de rejeições

De maneira similar ao cálculo de tempo de resolução, a contagem de número de rejeições se dá através da observação dos eventos de mudança de status. Uma rejeição é contabilizada toda vez que observa-se uma alteração de status *Em validação* para *Em progresso*.

5 RESULTADOS

Neste capítulo são apresentados os resultados das análises propostas na Seção 4.1

5.1 Assertividade vs. Métricas de Produtividade e Qualidade

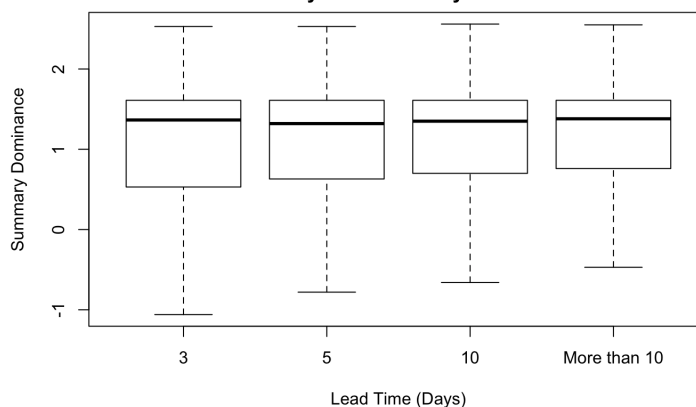
Para verificar a existência de relação entre assertividade e métricas de produtividade e qualidade, foram utilizadas medidas de dominância contidas no título e descrição dos cartões, dado que estes são os atributos textuais do cartão que carregam informação sobre a tarefa a ser desenvolvida e podem, portanto, conter expressões de emoção sobre a mesma. Conforme discutido na Seção 4.3.2, foi utilizada a função de agregação por maior distância da média para cálculo de sentimento.

5.1.1 Produtividade

A assertividade na descrição de um cartão influencia o tempo necessário para sua resolução?

A fim de visualizar o comportamento da dominância em função do tempo de resolução, as observações presentes no conjunto de dados foram divididas em grupos correspondentes a cartões resolvidos em até 3, 5, 10 dias e mais de 10 dias. As figuras 5.1 e 5.2 mostram o resultado. Observa-se que tanto o título quanto a descrição usam palavras cuja dominância é superior à dominância média do dicionário ANEW.

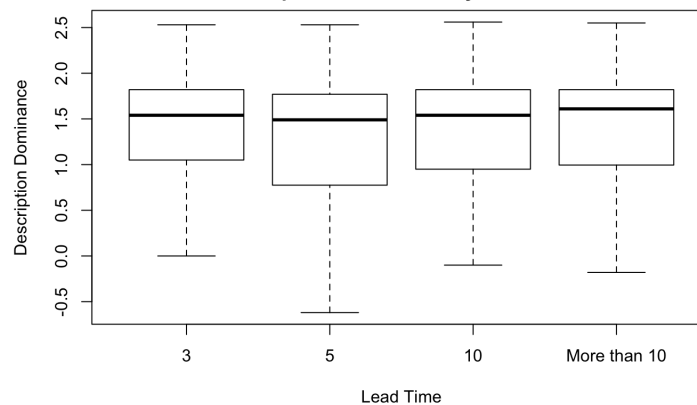
Figura 5.1: Dominância do título vs. tempo de resolução
Summary Dominance by Lead Time



A Figura 5.1 detalha a distribuição da dominância no título dos cartões para cada

grupo. Observa-se que a distribuição de dominância no título do cartão para os diferentes grupos de tempo de resolução é muito semelhante, sugerindo que o comportamento da variável dominância do título não se altera para os diferentes grupos de tempo de resolução. Calculando o coeficiente de correlação de Pearson entre a dominância do título e o tempo de resolução para o conjunto completo de cartões, obtemos o valor de $\rho = 0.02569012$, confirmando que a correlação é inexistente.

Figura 5.2: Dominância da descrição vs. tempo de resolução
Description Dominance by Lead Time



Na Figura 5.2 observa-se que a mediana para dominância da descrição é semelhante para todos os grupos e, embora exista variação entre os limites do primeiro e segundo quartil, ela não segue um padrão. O coeficiente de correlação de Pearson entre a dominância da descrição e o tempo de resolução para o conjunto completo de cartões, temos o valor de $\rho = 0.02776173$, evidenciando correlação inexistente.

Conclui-se que a assertividade contida no texto de um cartão, medida na forma de dominância, não tem relação com o tempo necessário para resolução de uma tarefa. Uma possível explicação para ausência de relação observável a partir dos dados é de que diversos fatores além da comunicação documentada no Jira podem ter impacto no tempo de resolução de um cartão.

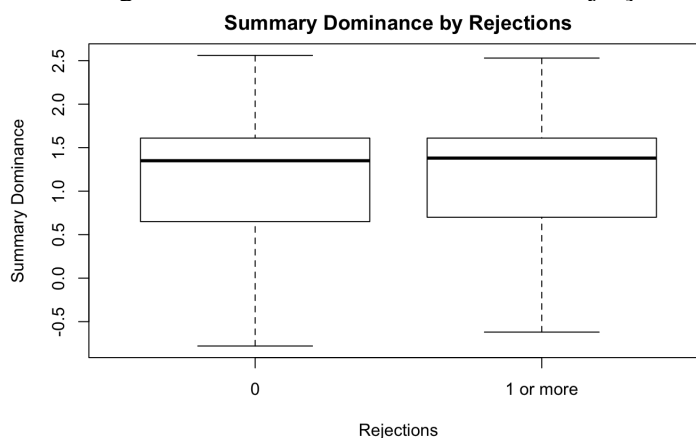
5.1.2 Qualidade

A assertividade na descrição de um cartão influencia o número de rejeições na etapa de validação?

Para medida de qualidade associada a um cartão, foi utilizada o número de rejeições na etapa de validação, onde um número menor de rejeições representa maior qualidade. Os cartões foram divididos em dois grupos representando tarefas que não foram

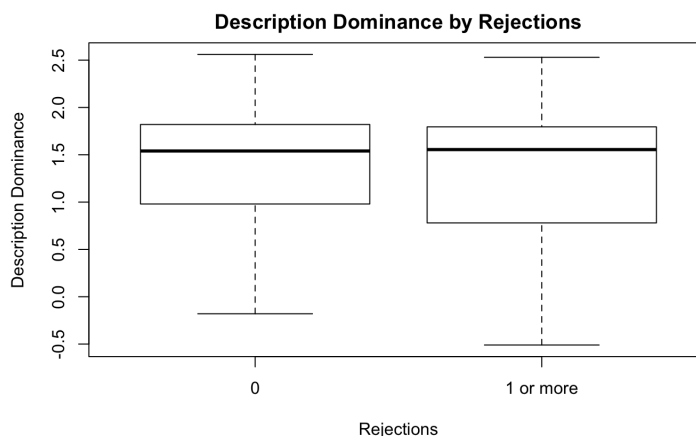
rejeitadas e tarefas que foram rejeitadas uma ou mais vezes durante a validação.

Figura 5.3: Dominância do título vs. rejeições



A Figura 5.3 mostra distribuições semelhantes de dominância no título dos cartões para ambos grupos. Ao aplicar um teste T não pareado com $\alpha = 0,05$ obtemos um valor $p = 0.7489$, mostrando que não existe diferença estatisticamente significativa entre as médias de dominância do título dos cartões para os dois grupos.

Figura 5.4: Dominância da descrição vs. rejeições



Na Figura 5.4 percebe-se uma pequena diferença entre os limites superiores dos quartis 1 e 2 dos grupos para dominância na descrição dos cartões. Aplicando um teste T não pareado com $\alpha = 0,05$ obtemos um valor $p = 0.3412$, mostrando que não existe diferença estatisticamente significativa entre as médias de dominância da descrição dos cartões para os dois grupos.

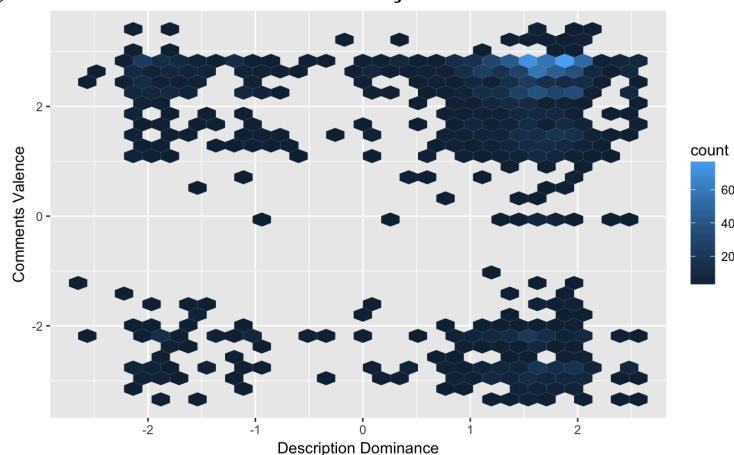
Assim podemos concluir que a assertividade no título e descrição de um cartão, representada pela medida de dominância não tem impacto no número de rejeições de um cartão na etapa de validação.

5.2 Assertividade vs. Sentimento nos Comentários

Existe relação entre assertividade na descrição de um cartão e o sentimento expresso em seus comentários?

A fim de entender o impacto da assertividade do cartão no sentimento das pessoas envolvidas na resolução da tarefa, foram analisadas as relações entre a dominância da descrição e as avaliações de emoção contidas nos comentários. A avaliação de emoção representativa para os comentários de um cartão consiste dos valores mais altos para cada uma das dimensões do modelo VAD encontrados nos comentários do mesmo. Conforme discutido na Seção 4.3.2, foi utilizada a função de agregação por maior distância da média para cálculo de sentimento presentes tanto na descrição quanto nos comentários.

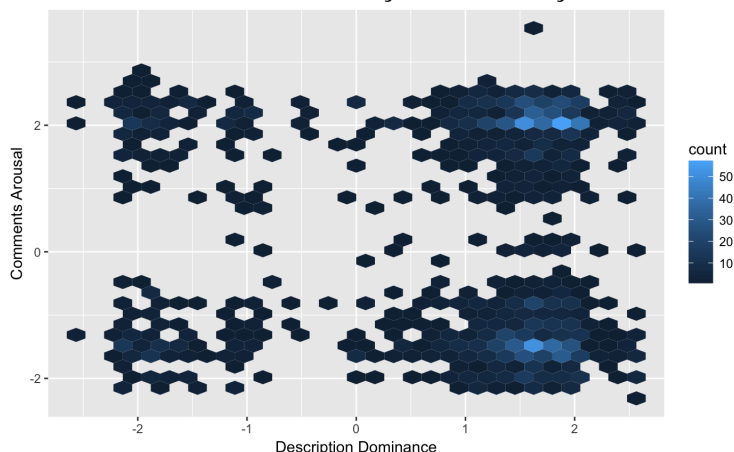
Figura 5.5: Dominância da descrição vs. valência dos comentários



Na Figura 5.5, podemos perceber que as observações dividem-se em quadrantes no plano de dominância vs. valência. O quadrante superior direito apresenta a maior concentração de ocorrências, representando uma possível relação entre altos valores de dominância e altos valores de valência. Tal observação sugere que cartões com descrição assertiva geram maior interesse e satisfação para pessoas envolvidas em sua resolução. O coeficiente de correlação de Pearson entre as variáveis, entretanto, tem valor de $\rho = 0.05140001$, indicando que a correlação inexistente entre dominância da descrição e valência nos comentários é inexistente.

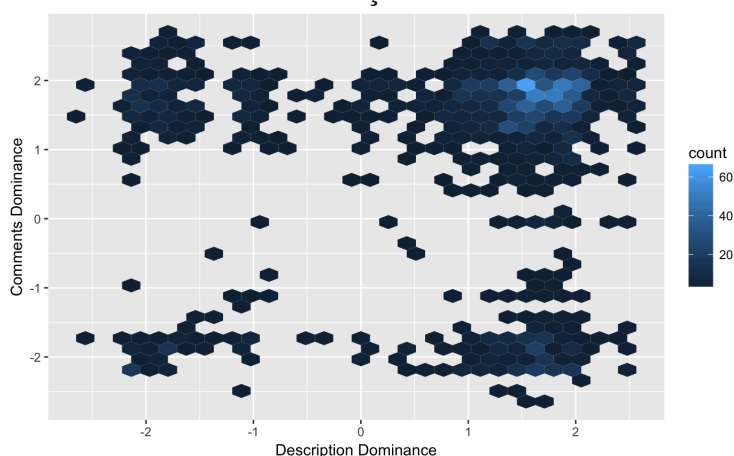
Na Figura 5.6, observa-se que a distribuição de excitação se dá de maneira balanceada entre avaliações altas e baixas, representando pouca interação entre as variáveis. Esta distribuição sugere que a assertividade da descrição não tem impacto no nível de agitação e associado a um cartão. Uma possível explicação seria de que o nível de excitação é mais facilmente influenciado por fatores externos à análise, como urgência para entrega

Figura 5.6: Dominância da descrição vs. excitação dos comentários



da tarefa. O coeficiente de correlação de Pearson entre as variáveis é de $\rho = 0.02058494$, mostrando que não existe correlação entre dominância da descrição e excitação dos comentários.

Figura 5.7: Dominância da descrição vs. dominância dos comentários



A Figura 5.7, apresenta que grande concentração de ocorrências no quadrante correspondente a valores altos de dominância da descrição e dos comentários, o que indica uma possível relação entre as variáveis. A interpretação atribuída a esta observação é de que a assertividade contida na descrição gera confiança nas pessoas envolvidas na resolução do cartão, refletindo-se na dominância dos comentários. O coeficiente de correlação de Pearson entre as variáveis dominância da descrição e dominância dos comentários é, no entanto, de $\rho = 0.1041633$, apontando que não existe correlação entre as variáveis.

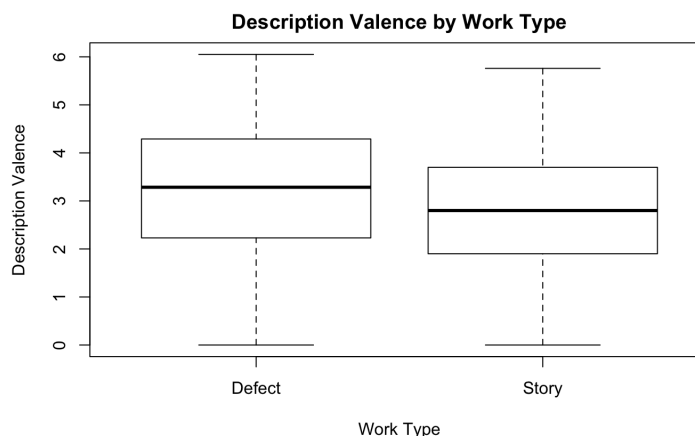
Enquanto a distribuição das avaliações de emoção dos comentários para dimensões de valência e dominância em função da dominância da descrição sugere a existência de relação, o cálculo de correlação de Pearson indica que as variáveis são independentes. Assim não é possível provar a existência de correlação linear entre as variáveis.

5.3 Tipo de trabalho vs. Sentimento no Cartão

Existe variação de sentimento na descrição do cartão em função do tipo de trabalho (nova funcionalidade vs. defeito)?

Foi avaliada a presença de sentimento em função do tipo de trabalho capturado pelo cartão. Somente cartões do tipo *nova funcionalidade* e *defeito* foram considerados para esta análise, uma vez que representam cerca de 93% das observações presentes no conjunto de dados. Para responder à pergunta, foi verificada a diferença entre as médias de avaliação de emoção da descrição do cartão para os dois tipos de cartão. Conforme discutido na Seção 4.3.1, foi utilizada a função de agregação por amplitude máxima para cálculo de sentimento.

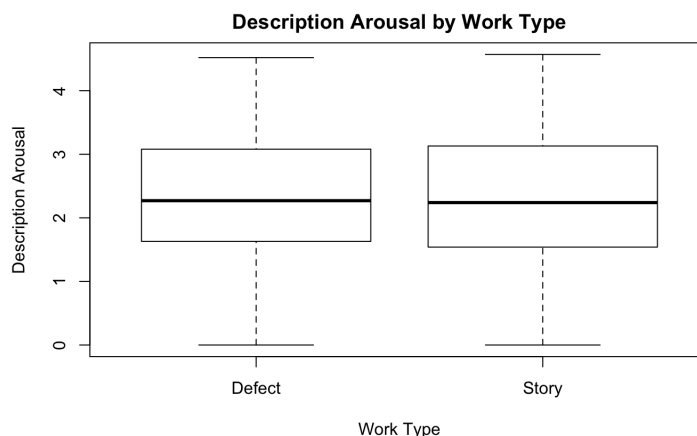
Figura 5.8: Valência da descrição vs. tipo de trabalho



Na figura 5.8 podemos observar uma diferença na distribuição da valência da descrição do cartão para os dois tipos de trabalho, sugerindo que a quantidade de emoção expressa na dimensão de valência é maior para cartões do tipo defeito. Aplicando um teste T não pareado com $\alpha = 0,05$ obtemos um valor $p = 1.003 \times 10^{-9}$, mostrando que a diferença entre as médias para os dois tipos de trabalho é estatisticamente significativa. Uma possível explicação para a diferença entre os grupos é de que cartões do tipo defeito frequentemente contém expressões negativas de sentimento, causando o aumento na quantidade de emoção para a dimensão de valência.

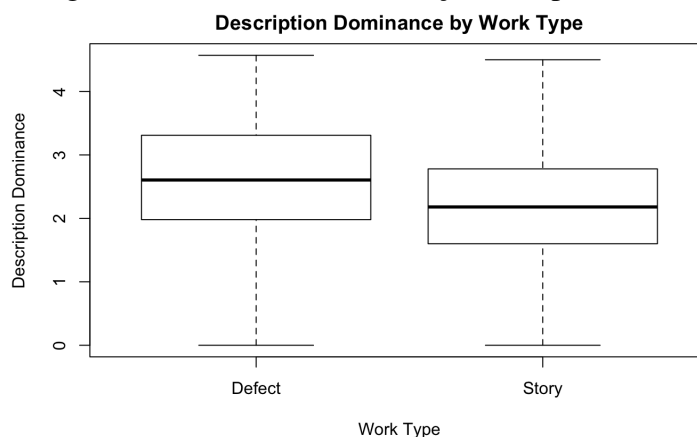
A figura 5.9 mostra que distribuição da excitação na descrição dos cartões é semelhante para os dois grupos, sugerindo inexistência de relação entre o tipo de trabalho e a quantidade e excitação presente na descrição. O teste T não pareado com $\alpha = 0,05$ produz um valor de $p = 0.4679$, confirmando que a diferença entre as médias para os dois tipos de trabalho não é significativa. Pode-se assumir que o contexto corporativo onde

Figura 5.9: Valência da descrição vs. tipo de trabalho



os cartões foram coletados não favorece demonstrações emocionais de agitação, portanto não se pode observar diferença entre os grupos.

Figura 5.10: Valência da descrição vs. tipo de trabalho



A figura 5.10 mostra diferença na distribuição da dominância da descrição do cartão para os dois tipos de trabalho, apontando para valores maiores de dominância na descrição de cartões do tipo defeito. Aplicando um teste T não pareado com $\alpha = 0,05$ obtemos um valor $p = 2.2 \times 10^{-16}$, mostrando que a diferença entre as médias dos dois grupos é significativa. A diferença entre os tipos de trabalho seria explicada pelo fato de que descrições de defeitos são frequentemente mais precisas que descrições de novas funcionalidades. Além de partir de uma seção de código já escrita, portanto conhecida pelo time, é comum que a descrição de um defeito inclua instruções para reprodução do mesmo, aumentando a assertividade.

Partindo destas análises podemos concluir que existe relação entre natureza do trabalho sendo reportado e a emoção da pessoa responsável pela criação do cartão. Tal relação é aparente para dimensões de valência e dominância, mas não observável para

dimensão de excitação.

6 CONCLUSÃO

Neste trabalho foi realizada uma análise das relações entre o sentimento expresso no enunciado de cartões Jira, coletados dentro do contexto do projeto Horse da organização Bay Bridge, e medidas de produtividade, qualidade, sentimento das pessoas envolvidas na resolução do cartão e tipo de trabalho representado pelo cartão. Foi concluído que:

- Não é possível observar relação entre a assertividade na descrição, representada pela dimensão de dominância do modelo VAD, e métricas de produtividade e qualidade da solução desenvolvida.
- A distribuição de emoção dos comentários para as dimensões de valência e dominância, sugere uma relação com a assertividade da descrição do cartão. Esta relação, no entanto, não seria uma relação linear, assim não foi possível confirmá-la através dos métodos de cálculo de correlação adotados.
- É possível observar diferença entre cartões representando novas funcionalidades de defeitos no que diz respeito a expressão de valência e dominância na descrição.

É possível abordar uma série de pontos para melhorar o resultado da análise em trabalhos futuros. O primeiro ponto a ser considerado é o número de projetos analisados. Uma vez que este trabalho considerou dados de um projeto apenas, os resultados refletem o contexto do mesmo, podendo diferir para times que adotem um conjunto diferente de práticas. Para tanto seria necessário incluir etapas de normalização de dados, visto que diferentes projetos podem ter usos distintos da ferramenta Jira.

O segundo ponto é o uso de um léxico específico para o domínio em questão. Como neste trabalho foi utilizado um léxico de propósito geral, é plausível assumir que diversas expressões com interpretações particulares dentro do domínio de desenvolvimento de software, ou até mesmo dentro da organização Bay Bridge, não foram associadas a semântica mais adequada.

O último ponto destacado é a exploração de diferentes maneiras de cálculo de sentimento e verificação da existência de relações. As funções de agregação utilizadas neste trabalho não consideram aspectos da estrutura sintática do texto, o que poderia representar um ganho na precisão das medidas de emoção. Além disso, as observações sugerem existência de relações não lineares, portanto o uso de métodos alternativos para verificação de tais relações poderia melhorar o resultado das análises.

REFERÊNCIAS

BACCIANELLA, S.; ESULI, A.; SEBASTIANI, F. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In: **in Proc. of LREC**. [S.l.: s.n.], 2010.

BECKER, K. Introdução à mineração de opiniões: Conceitos, aplicações e desafios. In: **Jornadas de Atualização em Informática - Congresso da SBC**. [S.l.: s.n.], 2014. p. 125–174.

BRADLEY, M. M.; LANG, P. J. **Affective Norms for English Words (ANEW): Instruction manual and affective ratings**. 1999.

DANESCU-NICULESCU-MIZIL, C. et al. A computational approach to politeness with application to social factors. **CoRR**, abs/1306.6078, 2013.

FELLBAUM, C. **WordNet: An Electronic Lexical Database**. [S.l.]: Bradford Books, 1998.

GRAZIOTIN, D. et al. Unhappy developers: Bad for themselves, bad for process, and bad for software product. **CoRR**, abs/1701.02952, 2017. Available from Internet: <<http://arxiv.org/abs/1701.02952>>.

GRAZIOTIN, D.; WANG, X.; ABRAHAMSSON, P. Understanding the affect of developers: Theoretical background and guidelines for psychoempirical software engineering. In: **Proceedings of the 7th International Workshop on Social Software Engineering**. New York, NY, USA: ACM, 2015. (SSE 2015), p. 25–32. ISBN 978-1-4503-3818-9. Available from Internet: <<http://doi.acm.org/10.1145/2804381.2804386>>.

KOLCHYNA, O. et al. Twitter sentiment analysis. **CoRR**, abs/1507.00955, 2015.

LIU, B. Sentiment analysis and subjectivity. In: **Handbook of Natural Language Processing, Second Edition**. Taylor and Francis Group, Boca. [S.l.: s.n.], 2010.

MÄNTYLÄ, M. et al. Mining valence, arousal, and dominance: Possibilities for detecting burnout and productivity? In: **Proceedings of the 13th International Conference on Mining Software Repositories**. New York, NY, USA: ACM, 2016. (MSR '16), p. 247–258. ISBN 978-1-4503-4186-8. Available from Internet: <<http://doi.acm.org/10.1145/2901739.2901752>>.

MEHRABIAN, A. Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament. v. 14, p. 261–292, Dez 1996.

MOHAMMAD, S. M.; TURNEY, P. D. Crowdsourcing a word-emotion association lexicon. **Computational Intelligence**, v. 29, n. 3, p. 436–465, 2013.

ORTU, M. et al. Are bullies more productive?: Empirical study of affectiveness vs. issue fixing time. In: **Proceedings of the 12th Working Conference on Mining Software Repositories**. Piscataway, NJ, USA: IEEE Press, 2015. (MSR '15), p. 303–313. ISBN 978-0-7695-5594-2. Available from Internet: <<http://dl.acm.org/citation.cfm?id=2820518.2820555>>.

RUBIN, K. S. **Essential Scrum: A Practical Guide to the Most Popular Agile Process**. 1st. ed. [S.l.]: Addison-Wesley Professional, 2012. ISBN 0137043295, 9780137043293.

SCRUM. 2018. <<https://www.scrum.org/resources/what-is-scrum>>. Acessado em 24/06/2018.

THELWALL, M. et al. Sentiment strength detection in short informal text. **J. Am. Soc. Inf. Sci. Technol.**, John Wiley & Sons, Inc., New York, NY, USA, v. 61, n. 12, p. 2544–2558, Dez 2010. ISSN 1532-2882. Available from Internet: <<https://doi.org/10.1002/asi.v61:12>>.

TSYTSARAU, M.; PALPANAS, T. Survey on mining subjective data on the web. **Data Mining and Knowledge Discovery**, Kluwer Academic Publishers, Hingham, MA, USA, v. 24, n. 3, p. 478–514, Mai 2012. Available from Internet: <<http://dx.doi.org/10.1007/s10618-011-0238-6>>.

WARRINER, A. B.; KUPERMAN, V.; BRYSSBAERT, M. Norms of valence, arousal, and dominance for 13,915 english lemmas. **Behavior Research Methods**, v. 45, n. 4, p. 1191–1207, Dez 2013.